SEMIANNUAL STATUS REPORT

NASA Grant NsG-354/05-003-016

16 November 1964 through 30 June 1965

ELECTRONICS RESEARCH LABORATORY

University of California, Berkeley

FOREWORD

The material presented here is in response to reporting requirements for NASA Grant NsG-354/05-003-016. The project is entitled, "Advanced Theoretical and Experimental Studies in Automatic Control and Information Systems." The principal investigators are Professors C. A. Desoer, E. Polak, D. Sakrison, and L. Zadeh.

This material will appear in the first consolidated semiannual progress report of the Electronics Research Laboratory, University of California, Berkeley, which is for the period of 16 November 1964 through 30 June 1965. The next E R L semiannual report will be for the period ending 31 December 1965.

ESTIMATIONS AND CONTROL OF NONLINEAR PHYSICAL SYSTEM

NASA Grant NSG-354 (Supplement 2)
T.C. Gaw (Prof. M. Aoki)

The problem under investigation is a class of adaptive control systems. It is assumed that the system dynamics is governed by

$$\frac{dx}{dt} = f(x,u,t),$$

where x = state variables, and

u = control.

The output or observation $y = h(x) +$ noise is assumed to be contaminated with noise of unknown probability distribution and the control u satisfies the equation

$$\frac{du}{dt} = g(u,t,y).$$

The purpose of this project is to find the best estimate of the state variables $x(t)$ under the circumstances given above. Using the criterion

$$F[u(t), \hat{x}(t), t] = \int_0^t dt \left\{ [k_1 \hat{x}^2(t) + k_2 u^2(t)] + 2[y(t) - h\hat{x}(t)]^2 \right\},$$

and invoking the invariant imbedding technique, the following results are obtained

$$\frac{d\partial}{dt} \simeq \partial[2f_s + \frac{2k_2 c}{k_1 s - \lambda(y-h(s)dh/ds}]$$
$$-\partial^2[k_1 - 2\lambda[y(t) - h(s)] \frac{d^2h(s)}{ds^2} + \lambda \frac{dh(s)}{ds}], \quad (1)$$

$$\frac{ds}{dt} = f - \partial[k_1 s = \lambda(y - h(s)) \frac{dh}{ds}], \quad (2)$$

where $x = s$ (present estimate),

$u = c$,

$f_s = \frac{\partial f}{\partial s}$.

To test the applicability of the above technique, a few computer runs will be made. The solution of Eq. 2 will be compared with the solution of the state variables obtained from known noise characteristics.

GAIN FUNCTION CHARACTERIZATION OF SYSTEMS

NASA Grant NSG-354 (Supplement 2)
D. Chazan (Prof. C.A. Desoer)

In the study of control systems it is almost universally assumed that the behavior of a (continuous time) control system is described by a differential equation partial or otherwise. In this work a different and more general description of control problems has been investigated which does not involve any differential euqations and attempts to reduce the problem to essentials. An inquiry was made into the possibility of translating some of the results of control theory into this setting. The answer obtained was in the positive at least in the special case that was studied.

At the base of this approach lies the concept of a gain function which assigns to every pair of states in the state-fine space the gain incurred in going from one state to the other. Such gain functions satisfy the following semi group condition:

$$C_{t_1t_2}(x,y) = \sup [C_{t_1t_3}(x,z) + C_{t_3t_2}(z,y)] \tag{1}$$

whenever $t_1 \leq t_3 \leq t_2$ (see quarterly report No. 14), and may be used to define a generalized system which is not necessarily describable by a differential equation. Thus a valid and mathematically interesting problem is the characterization of functions satisfying Eq. 1. A complete solution to this problem would certainly produce a generalized maximum principle as a by-product. The following is a first step in this direction:

<u>Theorem</u>: Let X be a locally convex linear topological space. Let T_t be a one parameter time group of continuous linear transformation on X which is continuous in t is some well defined sense. If $C_{t_1t_2}(x, y)$ is a gain junction (i.e. it satisfies Eq. 1), $C_{t_1t_2}(x, y) = C_{0,t_2-t_1}(0, y - T_{t_2-t_1}x)$, $C_{0,t}(0,x)$ is upper semi continuous in x and whenever $t_n \to 0$ $C_{0,t_n}(0,x_n) \to -\infty$ unless $x_n \to 0$ then the function $C_{0,t}(\cdot,\cdot)$ is convex. Furthermore it is possible to obtain a complete characterization of $C_{0,t}(0,\cdot)$ which in the special case when X is a Banach space and T_t is continuous in t uniformly on X reduces to:

$$C_{0,t}(0,x) = \sup\left\{ \int_0^t c(u(t)\,dt : \dot{x} = Ax + u, x(0) = 0, x(t) = 0 \right\}$$

for some convex function $c(\cdot)$ on X and a continuous linear operator A.

STABILITY OF NONLINEAR SINGLE-LOOP FEEDBACK SYSTEMS

NASA Grant NSG-354 (Supplement 2)
C.T. Lee (Prof. C.A. Desoer)

We consider the system S shown in Fig. 1

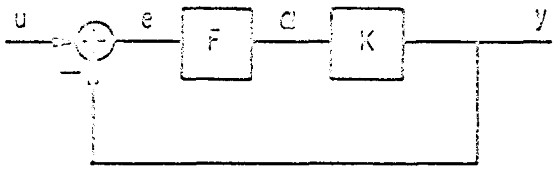


Fig. 1. The system S .

It is a single-input single-output feedback system. The system is characterized by following operator equations.

$$\begin{aligned} e &= u - y \ , \\ \alpha &= Fe \ , \\ y &= K\alpha \ , \end{aligned} \tag{1}$$

where F and K are operators.

(1) may be written as

$$(I + KF)\, e = u \tag{2}$$

$$y = KFe \, . \tag{3}$$

We shall introduce some notations before proceeding to the problems

1. H : a real Hilbert space, $<\cdot\,,\,\cdot>$ will denote the scalar product in H, and $\|\cdot\|$ corresponding norm.

2. Let x be a real valued function defined on $[0, \infty)$. Let us call P_T the projection operator such that

$$\begin{aligned} (P_T x)(t) &= x(t) \qquad \text{for } 0 \le t \le T \\ &= 0 \qquad \text{elsewhere} \, . \end{aligned}$$

3. H_e : an extension of H such that

   a) the elements of H_e are functions defined on $[0, \infty)$

b) if $x \in H_e$

$P_T x \in H$ for all finite T.

4. An operator $F: H_e \longrightarrow H_e$ is said to be of a finite gain if there are two real numbers $g(F)$ and r_F such that

$\|P_T Fx\| \leq g(F) \|P_T x\| + r_F$ for all $T \geq 0$ and for all $x \in H_e$.

5. L^2 : a set of square integrable real valued functions on $[0, \infty)$.

6. L_e^2 : a set of locally square integrable real valued functions on $[0, \infty)$. L_e^2 is an extension of L^2.

7. An operator F is said to be nonanticipative if and only if

$$P_T F = P_T F P_T \qquad \text{for all } T \geq 0$$

8. We shall often write x_T for $P_T x$, and we adopt the following conventions

$$\|x\|_T \triangleq \|x_T\| \qquad \forall\, x \in H_e,$$

$$< x, y >_T \triangleq < x_T, y_T > \qquad \forall\, x, y \in H_e$$

The main results obtained shall be stated as following theorems

Theorem 1

Consider the operator Eq. 2

Suppose that

i) all solutions of (2) for $u \in H$ belong to H_e

ii) $F: H_e \longrightarrow H_e$ and $K: H_e \longrightarrow H_e$

iii) there are sequences of real numbers

$$\{k_i^{(1)}\}_{i=0}^{i=n}, \ \{k_i^{(2)}\}_{i=0}^{i=n}, \ \{k_i^{(3)}\}_{i=0}^{i=n}, \ \{k_i^{(4)}\}_{i=0}^{i=n}, \ \{P_i\}_{i=0}^{i=n}$$

such that for all $x \in H_e$ and for all $T \geq 0$

$$< x, Fx >_T \geq \sum_{i=0}^{n} \{k_i^{(1)} \|x\|_T^{P_i} + k_i^{(2)} \|Fx\|_T^{P_i}\}$$

and

$$< Fx, KFx >_T \geq \sum_{i=0}^{n} \{k_i^{(3)} \|Fx\|_T^{P_i} + k_i^{(4)} \|KFx\|_T^{P_i}$$

where

$$0 \leq P_0 < P_1 < \cdots < P_n$$

$$1 < P_n .$$

Under the above assumptions all solutions of (2) for $u \in H$ belong to H provided that one of the following two conditions is satisfied.

a. $k_n^{(1)} + k_n^{(4)} > 0$ and $k_n^{(2)} + k_n^{(3)} \geq 0$, F is of finite gain.

b. $k_n^{(1)} + k_n^{(4)} \geq 0$ and $k_n^{(2)} + k_n^{(3)} > 0$, $K: H \rightarrow H$.

Many researchers [1], [2] have investigated the stability of the nonlinear system S. Theorem 1 is more general than those obtained by Sandberg [1] and Zames [2] under somewhat less restrictive assumptions.

The next result is a special class of nonlinear system shown in Fig. 1.

We shall make the following assumptions:

Assumptions

F1. The operator F is characterized by

$$(Fe)(t) = \alpha(t) = \varphi[e(t)],$$

where the nonlinear characteristic φ is a piecewise continuous function defined on $(-\infty, \infty)$.
Furthermore,

F2. $\quad 0 \leq e\varphi(e) \leq ke^2 \quad$ for all $-\infty < e < \infty$,

and $\quad \varphi(0) = 0$.

Fe. $\quad \int_0^\sigma \varphi(\xi)\, d\xi \longrightarrow \infty \quad$ as $|\sigma| \longrightarrow \infty$.

K1. The subsystem K is characterized by its input-output relation

$$y = z + k\alpha,$$

where $z(\cdot)$ is the zero input response which depends on the state of K at time 0, and K is a nonanticipative operator mapping from L^2 into itself.

---

[1] I.W. Sandberg, "Some Results on the Theory of Physical Systems Governed by Nonlinear Functional Equations," Bell System Technical Journal, Vol. XLIV, May 1965 p. 871-898.

[2] G. Zames, "On the Stability of Nonlinear Time-Varying Feedback Systems," NEC Proceeding Sept., 1964 p. 725-730.

K2. The operator DK is also assumed to be a nonanticipative operator mapping from L^2 into itself where $(DKx)(t) \triangleq \frac{d}{dt}(Kx)(t)$.

K3. For all initial states, $z(\cdot)$ and $\dot{z}(\cdot)$ are in L^2, and $z(0)$ is finite.

K4. The zero input response of the feedback system S for all initial states is assumed to be locally square integrable, i.e., $y(\cdot) \in L_e^2$ for all $z(\cdot)$ satisfying K3.

Remark

In fact K1, K2, and K3 imply K4 because $z(\cdot)$ and $k\alpha$ are differentiable and hence $y(\cdot)$ is continuous.

Theorem 2

Consider the system S shown in Fig. 1. Let $u = 0$. Suppose that the assumptions F1 - F3 and K1 - K4 are satisfied. Under these assumptions if there are positive numbers q and δ such that

$$< x, (K + q\, DK + \frac{1}{k})\, x > \geq \delta \|x\|^2 \quad \forall\, x \in L^2 ,$$

then the zero input response of the feedback system S is bounded square integrable functions for all initial states, i.e.,

$$y(\cdot) \in L^{\infty} \cap L^2 .$$

Theorem 2 is an extension of Popov's Theorem for stability. In refs. [3] and [4] the subsystem K is assumed to be a linear time invariant system. However, in Theorem 2 no assumption is made concerning linearity or time invariance.

Part of Theorem 2 has been reported and proved [5]. The significance of this result is the fact that it is derived without using Liapunov's functions. Asympotic stability, i.e., $y(t) \to 0$ as $t \to \infty$, can be also obtained by making additional assumptions on the subsystem K. A slightly more general result has been obtained.

---

[3] M. Aizerman and F.R. Gantmacher, "Absolute Stability of Regulator Systems," (Translated from Russian by E. Polak) San Francisco, Holden Day, 1964.

[4] C.A. Desoer, "A Generalization of the Popov Criterion," IEEE Trans., AC-10, 2, p.182-184, April 1965.

[5] C.A. Desoer and C.T. Lee, "Stability of Single-Loop Feedback Systems," Notes on System Theory, VII Feb. 1965.

Theorem 3

Let $u = 0$. Suppose that the assumptions F1 - F3 and K1 - K4 are satisfied. Under these assumptions if there is a positive number δ and if there is a monotonically decreasing function $q(\cdot)$ such that

$$0 < \epsilon \leq q(t) \leq q(0) < \infty$$

and

$$< x, (K + q\,DK + \frac{1}{k})\, x > \geq \delta \,\|x\|^2 \qquad \forall x \in L^2$$

then, for each initial state, the zero input response of the feedback system S is bounded and square integrable function over $(0, \infty)$ i.e.,

$$y(\cdot) \in L^{\infty} \cap L^2 .$$

Remark

Theorem 2 is in fact a special case of Theorem 3, i.e., $q(t) = q = \text{const.}$

Since frequency domain analysis is not applicable to the general nonlinear system, all the results mentioned above are derived in time domain. Further extension of Theorem 2 to higher order systems is under investigation.

All the results obtained are being prepared for publication.

TIME OPTIMAL CONTROL OF A CLASS OF PULSE WIDTH MODULATED SAMPLED DATA SYSTEMS

NASA Grant NsG-354 (Supplement 2) and NSF Grant GP-2413
M. Canon and C.D. Cullum (Prof. E. Polak)

By combining work by M.D. Canon on the time optimal control of pulse amplitude modulated linear discrete systems* and the work of C.D. Cullum and E. Polak on the equivalence of optimal control problems† a computational algorithm has been developed for the optimal control of a class of pulse width modulated linear sampled data systems. The class of problems considered and the algorithm developed are described in the following.

*Described

† ERL Consolidated Quarterly Progress Report No. 15, Aug. 16-Nov. 15, 1964, p. VII-11

1. PWM Linear Sampled Data System

It is assumed that the plant is described by a set of linear differential equations of the form

$$\dot{\underline{x}} = \underline{A}\,\underline{x} + \underline{b}\,u$$

$$A = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \qquad \lambda_n < \lambda_{n-1} < \cdots < \lambda_1 < 0$$

$$\underline{b} = \text{col}\,(b_1, \ldots, b_n)$$

$$\underline{x} = \text{col}\,(x_1, \ldots, x_n)$$

$$u = \text{scalar}$$

The state of the system is assumed to be sampled once every second (on a normalized time scale), and the control function u(t) is given by

$$u(t) = \begin{cases} \text{sgn}\, u_n & n-1 \leq t < n-1+|u| \\ 0 & n-1+|u_n| \leq t < n \end{cases}$$

where the u_n $n = 1, 2, \ldots$ are constants satisfying $|u_n| \leq |$ for all n . The optimal control problem is stated in the following form:

Given that the system is in an initial state $\underline{x}_0$ at t = 0, find a control sequence $u_1, u_2, \ldots, u_N$ which brings the system to the origin in a number of sampling periods, N, which is less than or equal to the nunber of sampling periods required with any other control sequence.

The following definitions are used in the discussion of the algorithm:

$$\underline{w}_i(u_i) \triangleq e^{-(i-1)\underline{A}}\left[e^{-\underline{A}|u_i|} - \underline{I}\right]\underline{b}$$

$$\underline{r}_i \triangleq \underline{w}_i(1) = e^{-(i-1)\underline{A}}\left[e^{-\underline{A}} - \underline{I}\right]\underline{b}$$

$$\text{sat}(x) \triangleq \begin{cases} x & \text{if}\, |x| \leq 1 \\ \text{sgn}\, x & \text{if}\, |x| > 1 \end{cases}$$

$$R_N \triangleq \{\underline{x}_0 / \underline{x}_0 = -\sum_{i=1}^{N} \underline{w}_i(u_i)\,\text{sgn}\,u_i,\ |u_i| \leq 1,\ i=1, \ldots, N\}$$

= the set of states from which the origin is reachable in N samplings periods or less.

2. Optimal Control Algorithm

The algorithm is based on the following conjectured theorem.

Theorem: Under the conditions of the problem stated above, the sets $C_N \subset E^n$, $N = n, n+1, \ldots$ defined by

$C_N = \{ \underline{c} / \text{the set of indices } \{ i/i \leq N, |< \underline{r}_i, \underline{c} >| \leq 1\} \text{ has cardinality } \geq n \}$ have the properties that

(a) C_N is homeomorphic to R_N for every $N \geq n$ with the homeomorphism $f: C_N \rightarrow R_N$ given by

$$\underline{f}(\underline{c}) = - \sum_{i=1}^{N} \underline{w}_i \, (\text{sat} < \underline{n}, \ \underline{c} >) \ \text{sgn} < \underline{r}_i, \underline{c} >$$

(b) for any $\underline{x}_0 \in R_N$ there exists a control sequence $u_1, u_2, \ldots, u_N$ which brings the system from the initial state $\underline{x}_0$ to the origin in N sampling periods, which is given by $u_i = \text{sat} < \underline{r}_i, \underline{f}^{-1}(\underline{x}_0) > \quad i = 1, \ldots, N$

(c) $\underline{f}$ and $\underline{f}^{-1}$ are piecewise differentiable. Some remaining details of the proof of this theorem are being worked out and it is expected that the material will be submitted for publication in the near future.

Given the initial state, $\underline{x}_0$, of the system, the basic procedure used in this algorithm is to examine successively R_n, $R_{n+1}, \ldots$ to determine whether $\underline{x}_0$ belongs to these sets, stopping at the first N for which $\underline{x}_0 \in R_N$. This N represents the minimum number of sampling periods to bring the system to the origin. To determine whether $\underline{x}_0$ is in R_N for a fixed N, the algorithm seeks a solution of the equation

$$\underline{f}(\underline{c}) = \underline{x}_0$$

with $\underline{c} \in C_N$. This is accomplished using standard techniques for the solution of a set of nonlinear equations. If a solution is not found, N is increased by one and the procedure is repeated. If a solution, $\underline{c}_0 \in C_N$, is found, then an optimal control is given by

$$u_i = \text{sat} < \underline{r}_i, \underline{c}_0 > \qquad i = 1, \ldots, N.$$

The procedure described here has not been tested in practice. However, an equivalent procedure involving a more complicated function f was examined using a second-order system with initial states less than or equal to 20 sampling periods from the origin. Typical computation times for initial states picked at random were of the order of

.4 to 5 seconds or less on an IBM 7094 computer. The authors expect to obtain considerable computational experience with this algorithm during the next few months. The relative simplicity of the present procedure, as compared with the one for which computation times were quoted, would lead one to suspect that these times could be reduced by a factor of two or more. The actual times obtained will be reported in the next progress report.

REAL-TIME IDENTIFICATION OF TRANSFER

NASA Grant NsG-354 (Supplement 2)
K.Y. Wong (Prof. E. Polak)

A new method of real-time identification of a transfer function under normal operation with noise disturbance is discussed. The problem is stated and a summary is given of the writer's previous result in using testing function to convert the problem of estimating the coefficients of a differential equation into one of estimating the coefficients of an algebraic equation. An outline is also given of the use of instrumental variables to obtain a consistent estimate of the desired parameters. It is shown that the solution can be re-written into a recursive form, enabling the estimates of the parameters to be up-dated. The optimal shape of the testing function with respect to a suitable criterion is also discussed.

1. Problem Statement

Assume that the input $u(t)$ and the output $V(t)$ of time invariant system obey the differential equation:

$$a_2 \frac{d^2V}{dt^2} + a_1 \frac{dV}{dt} + a_0 V = b_1 \frac{du}{dt} + u . \tag{1}$$

However, $V(t)$ cannot be observed, but $v(t)$ can be measured, where

$$v(t) = V(t) + e(t) ,$$

$e(t)$ is a stationary noise process with zero mean (not necessary white noise). We assume that the variables $v(t)$ and $u(t)$ have been observed over some interval of time which we then divided into subintervals of length T. The problem is to estimate the coefficients a_2, a_1, a_0, b_1 .

A second-order system is chosen for convenience of illustration; the theory that follows can be trivially extended to higher order systems.

Let $g(t)$ be a continuous function such that

$$g(0) = g^{(1)}(0) = g^{(2)}(0) = g(T) = g^{(1)}(t) = g^{(2)}(T) = 0.^{*} \tag{2}$$

The function $g(t)$ will be called a testing function. Multiplying both sides of Eq. 1 by $g(t)$ and integrating from zero to T, we get,

---

* $g^{(j)}(t)$ denotes the jth derivative of $g(t)$.

$$a_2\int_0^T g(t)\ V^{(2)}(t)\ dt + a_1\int_0^T g(t)\, V^{(1)}\ dt + \int_0^T g(t)\, V(t)\ dt$$

$$= b_1\int_0^T g(t)\, u^{(1)}(t)\ dt + \int_0^T g(t)\, u(t)\ dt\ . \tag{3}$$

Equation 3, after integration by parts, gives

$$a_2\int_0^T g^{(2)}(t)\, v(t)\ dt + a_1\int_0^T -g^{(1)}(t)\ dt + a_0\int_0^T g(t)v(t)\, dt + b_1\int_0^T g^{(1)}(t)a(t)\ dt$$

$$= \int_0^T g(t)\, u(t)\, dt + \sum_{j=0}^{2} (-1)^j a_j \int_0^T g^{(j)}(t)\ e(t)\ dt.$$

Note that the integrals

$$\int_0^T g^{(j)}(t)\ v(t)\ dt, \qquad j = 0, 1, 2,$$

and

$$\int_0^T g^{(j)}(t)\ u(t)\ dt, \qquad j = 0, 1,$$

can easily be computed and that

$$\sum_{j=0}^{2} (-1)^j a_j \int_0^T g^{(j)}(t)\ e(t)\ dt$$

is the disturbance quantity. The process of multiplying by a testing function and integrating from 0 to T can now be repeated on the observed data $v(t)$, $u(t)$, for t in the time intervals $[T, 2T]$, $[2T, 3T]$ .... and so on. These will be referred to as the first, 2nd, etc. intervals of integration. Let us define

$$x_{kj} \triangleq (-1)^j \int_0^T g^{(j)}(t)\ v(t)\ dt \qquad j = 0, 1, 2 \tag{4a}$$

$$x_{k3} \triangleq \int_0^T g^{(1)}(t)\ u(t)\ dt \tag{4b}$$

$$y_k \triangleq \int_0^T g(t)\, u(t)\, dt \tag{4c}$$

$$\epsilon_j \triangleq \sum_{j=0}^{2} (-1)^j a_j \int_0^T g^{(j)}(t)\, e(t)\, dt \tag{4d}$$

$$a_3 \triangleq b_1 \tag{4e}$$

The index k in Eq. 4 refers to the interval of the integration. The following set of algebraic equations is then obtained.

$$\begin{bmatrix} x_{10} & x_{11} & x_{12} & x_{13} \\ x_{20} & x_{21} & x_{22} & x_{23} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ x_{N0} & x_{N1} & x_{N2} & x_{N3} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ \vdots \\ Y_N \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \vdots \\ \vdots \\ \epsilon_N \end{bmatrix}.$$

In symbolic matrix notation, Eq. 5 can be written as

$$X_N \underline{a} = \underline{y}_N + \underline{\epsilon}_N\,. \tag{5}$$

Define

$$\underline{X}_k \triangleq \begin{bmatrix} X_{k0} \\ X_{k1} \\ X_{k2} \\ X_{k3} \end{bmatrix} \tag{6}$$

Note that $E(x_{kj}\, e_j) \neq 0$; $k = 1, 2, \ldots N$; $j = 0, 1, 2$ *

The problem of estimating the coefficients of the differential Eq. 1 is transformed into finding a statistical estimate of $\underline{a}$ in Eq. 5.

---

* "E" denotes expectation.

2. U[illegible] of Instrumental Variables to Obtain Consistent Estimates

Suppose we generate a vector instrumental stationary stochastic process $\{\underline{z}_k\}$ where

$$\underline{z}_k \triangleq \begin{bmatrix} z_{k0} \\ z_{k1} \\ z_{k2} \\ z_{k3} \end{bmatrix} \tag{7}$$

and k is the timing index, such that

(i) $E\left(\underline{z}_{Nj}\ \underline{\epsilon}_N^T\right) = 0, \qquad j = 0, 1, 2, 3,$

where $\underline{z}_{Nj}$ is a vector with elements $(z_{1j}, z_{2j}, \ldots z_{Nj})$, and $\underline{\epsilon}_N^T$ is the transpose of $\underline{\epsilon}_N$;

and (ii) $E\left(\underline{z}_{Nj}\ \underline{x}_{nj}\right) \neq 0, \qquad j = 0, 1, 2, 3,$ (8)

where $\underline{x}_{Nj}$ is the $\underline{j}$th column vector defined by the matrix X_N in Eq. 5.

Now use the estimator $\hat{a}_N = \left(Z_N^T\ X_N\right)^{-1} Z_N^T\ \underline{y}_N$ (9)

where

$$Z_N = \begin{bmatrix} z_{10} & z_{11} & z_{12} & z_{13} \\ z_{20} & z_{21} & z_{22} & z_{23} \\ \vdots & \vdots & \vdots & \vdots \\ z_{N0} & z_{N1} & z_{N2} & z_{N3} \end{bmatrix}$$

Therefore $\hat{a}_N = \underline{a} + \left(\frac{1}{N}\ Z_N^T\ X_N\right)^{-1} \left(\frac{1}{N}\ Z_N^T\ \underline{\epsilon}_N\right).$

By Slutsky's Theorem [1]

$$\operatorname*{plim}_{N \to \infty} \hat{a}_N = \underline{a} + \left[\operatorname{plim}\left(\frac{1}{N} Z_N^T\ X_N\right)\right]^{-1} \left[\operatorname{plim}\left(\frac{1}{N}\ Z_N^T\ \underline{\epsilon}_N\right)\right]\ *$$

[1] S. Wilks, Mathematical Statistics, John Wiley and Sons, Inc. New York, (1962), p. 102.

* "plim" denotes a limit in probability.

It can be shown that (a) $\underset{N \to \infty}{\text{plim}} \frac{1}{N} Z_N^T X_N = E\, \underline{z}_k\, \underline{x}_k^T$ (10)

$$\triangleq \Sigma_{zx}$$

where $\underline{x}_k$ was defined in Eq. 6, and it is assumed that the 4 x 4 square matrix has finite components; and

(b) $\text{plim}(\frac{1}{N} Z_N\, \underline{\epsilon}_N) = 0$;

therefore plim $\hat{a}_N = \underline{a}$. If the noise process e(t) is white, the asymptotic variance of a_N has been proved to be $\frac{\sigma^2}{N} \Sigma_{zx}^{-1} \Sigma_{zz} \Sigma_{xz}^{-1}$ (11)

where Σ_{zz}, Σ_{zx}, Σ_{xz} are 4 x 4 square matrices defined similarly as in Eq. 10 and $\sigma^2 = E\, \epsilon_k^2$.

If e(t) is not white, an asymptotic expression for the variance of a_N can still be obtained but it is more complicated than Eq. 11. The measurable output v(t) of Eq. 1 is correlated with u(t), but u(t) is independent of e(t); hence an obvious way to generate the instrumental variable process is to derive it from u(t).

3. Recursive Estimate

Let X_{N+1}, Z_{N+1}, Y_{N+1} be partitioned as follows:

$$X_{N+1} = \begin{bmatrix} x_{10} & x_{11} & x_{12} & x_{13} \\ x_{20} & x_{21} & x_{22} & x_{23} \\ \vdots & \vdots & \vdots & \vdots \\ x_{N0} & x_{N+1} & x_{N+1} & x_{N3} \\ \hline x_{N+1,0} & x_{N+1,1} & x_{N+1,2} & x_{N+1,3} \end{bmatrix} = \begin{bmatrix} X_N \\ \\ \hline m_{N+1} \end{bmatrix}$$

$$Z_{N+1} = \begin{bmatrix} Z_N \\ \hline q_{N+1} \end{bmatrix} , \qquad y_{N+1} = \begin{bmatrix} y_N \\ \hline y_{N+1} \end{bmatrix}$$

It has been shown that the estimator a_N in Eq. 9 obeys the following recursive relation:

$$\hat{a}_{N+1} = \hat{a}_N - \left[P_N q_{N+1}^T (1 + m_{N+1} P_N q_{N+1}^T)^{-1}\right] \left[m_{N+1} \hat{a}_N - Y_{N+1}\right] \quad (12)$$

and

$$P_{N+1} = P_N - P_N a_{N+1}^T (1 + m_{N+1} P_N a_{N+1}^T)^{-1} m_{N+1} P_N$$

with $P_N \triangleq (Z_N^T X_N)^{-1}$. Note that $(1 + m_{N+1} P_N q_{N+1}^T)$ is a scalar.

4. Optimal Shape of the Testing Function

For a functional of optimality of the testing functions, with T given, we chose

$$\rho(T) = \frac{\int_0^T \int_0^T R_s(s-t)\, g(t)\, g(s)\, dt\, ds}{\int_0^T \int_0^T R_n(s-t)\, g(t)\, g(s)\, dt\, ds}, \quad (13)$$

where $R_s(t)$ = autocorrelation function of V(t), and

$R_n(t)$ = autocorrelation function of e(t).

Both $R_s(t)$ and $R_n(t)$ are assumed to be known. We shall now show how to choose a continuous function g(t) that maximizes $\rho(T)$ subject to the boundary conditions of Eq. 2.

Case A: $R_n(t) = \delta(t)$, i.e., e(t) is a white-noise process. It is well known that g(t) maximizes the ratio (without boundary conditions)

$$\frac{\int_0^T \int_0^T R_s(s-t)\, g(t)\, g(s)\, dt\, ds}{\int_0^T g^2(s)\, ds} \quad (14)$$

if and only if it satisfies the following integral equation:

$$\int_0^T R_s(s-t)\, g(s)\, ds = \lambda_{max}\, g(t) \;, \tag{15}$$

where λ_{max} is the maximum eigenvalue of Eq. 15. If the maximization of the ratio in Eq. 14 is subject to h boundary conditions, then the maximum of the ratio in Eq. 14 $\rho(T)$, by the Courant mini-max lemma [2], satisfies the relation

$$\lambda_{h+1} \leq \lambda(T) \leq \max,$$

where λ_{h+1} is the (h+1)st largest eigenvalue of Eq. 15.

Solution of the integral equation (Eq. 15) by analytical methods is difficult in practice; hence, approximations in $L^2_{[0,T]}$ are used. Let

$$R_s(s-t) \approx \sum_{i=1}^{M} \sum_{j=1}^{M} \gamma_{ij}\; \psi_i(t)\; \psi_j(s) \;, \tag{16}$$

where

$$\gamma_{ij} \triangleq \int_0^T \int_0^T R_s(s-t)\; \psi_i(t)\; \psi_j(s)\; dt\, ds \;, \tag{17}$$

where $\{\psi_i(t)\}$ is a complete orthonormal basis for $L^2_{[0,T]}$. With this series expansion of $R_S(s-t)$, the finding of an approximation solution of Eq. 15 with or without constraints of Eq. 2 on g(t), leads to a matrix eigenvalue problem. If the Fourier transform of $R_s(t)$ is a ratio of polynomials in ω^2, and if $\{\psi_i(t)\}$ are chosen to be sine and cosine functions, then γ_{ij} can be computed readily by the residue method of complex variable theory.

It would be desirable to know $\rho(T)$ explicitly as a function of T, since then it would be possible to make a better choice of T; unfortunately $\rho(T)$ can only be computed for a range of T by repeatedly solving Eq. 15. On the other hand, upper and lower bounds for $\lambda_{max}(T)$ can be easily obtained as functions of T without solving the integral equation. It has been proved by Bellman and Letter [3] that

---

[2] R. Courant and D. Hilbert, Methods of Mathematical Physics, Interscience (1953), p. 31.

[3] R. Bellman and R. Latter, "On the integral equation $\lambda f(x) = \int_0^a K(x-y)\, dy$," Proc. Amer. Math. Soc., 3, pp. 884-891 (1952).

$$\frac{2}{T} \int_0^T (T-t)\, R_s(t)\, dt \leq \lambda_{max}(T) \max_{0 \leq x \leq T} \int_0^T R_s(x-y)\, dt, \qquad (18)$$

In particularly, if $R_s(t)$ is a function which decreases monotonically for $|t| \uparrow \infty$ the right hand side of the inequality (Eq. 18) reduces to

$$2 \int_0^{\frac{T}{2}} R_s(t)\, dt \, .$$

When the Fourier transform of $R_s(t)$ is a rational polynomial in ω^2, certain necessary conditions, as well as some sufficient conditions for the location of the zeros and poles of the Fourier transform of $R_s(t)$, have been developed to ensure the monotonicity of $R_s(t)$.

Case B: When $R_N(t) \neq \delta(t)$, the optimization problem in Eq. 13 can be solved by expanding $R_s(s-t)$ and $R_N(s-t)$ into series of sine and cosine terms, and the problem again reduces to a matrix eigenvalue problem.

5. Conclusions

The method of identification outlined in this report can be used in systems under normal operation condition and it can update the estimate easily when new data are received. Further work will be carried out choosing an optimal interval of integration and an optimal set of instrumental variables.

DECOMPOSITION OF LARGE SYSTEMS

NASA Grant NsG-354 (Supplement 2)
P. Varaiya (Prof. L.A. Zadeh)

A basic problem in nonlinear programming is the following:

$$\text{Maximize } \{ f(x) \mid g(x) \geq 0, \ x \geq 0 \} \qquad \text{NP}$$

where $x \in E^n$, $g: E^n \longrightarrow E^m$ is a differentiable mapping, and f is a real-valued, differentiable function. Necessary conditions for the solution of the NP problem were first formulated by Kuhn and Tucker [1]. Their result essentially consists of a non-trivial extension of the classical theory of Lagrange multipliers.

[1] H.W. Kuhn and A.W. Tucker, "Nonlinear Programming," Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Berkeley and Los Angeles, University of California Press,(1951), pp. 481-492 .

Some effort [2] has since been devoted to extend the K.T results to problems where the variables range over more general spaces or are subject to more general constraints. We have considered the following problem:

$$\text{Maximize } \{f(x) \mid g(x) \in \Omega_Y, \ x \in \Omega\} \qquad \text{P}$$

where X and Y are real Banach spaces, $x \in X$, $\Omega \subseteq Y$ is an arbitrary set and $\Omega_Y \subseteq Y$ is a closed convex set. $g: X \to Y$ and $f: Y \to$ Reals are arbitrary Frechet-differentiable functions. The necessary conditions which we obtain are very analogous to those of Kuhn and Tucker and are based on the well known geometric fact that two disjoint closed convex sets one of which is compact, can be strictly separated by a closed hyperplane. Such a general viewpoint also lends insight into some theoretical aspects of optimal control and shows the essential similarilty between the structure of problem P and those of optimal control. Thus the necessary conditions for the solution of problem P then gives us (local) maximum principles for both discrete -- and continuous -- time optimal control problem. Moreover, in the case of discrete time problems, we can assume that the state vector and the control belong to any real B-spaces. This allows us to obtain necessary conditions for the solution of stochastic discrete time optimal control problems.

Finally we show that the problem P is related to a saddle-value or game-theoretic problem. This viewpoint helps us to develop techniques for the solution by decomposition of some large nonlinear programming problems. Some of these decomposition results will appear in the Fall issue of the Control Section of SIAM Journal. Most of these results will be published soon as an ERL Report. In the near future we hope to obtain more algorithms for the practical solution of these problems and extend the theory to cases where we have a vector-valued or "minimax" type cost function.

A NEW ALGORITHM FOR A CLASS OF QUADRATIC PROGRAMMING PROBLEMS WITH APPLICATION TO CONTROL *

NASA Grant NsG-354 (Supplement 2)
M. Canon (Prof. E. Polak)

An algorithm is given which can be used to solve the following optimal control problems for linear discrete time systems: minimum energy, minimum time, and minimum energy plus time. Each of the optimal control problems is reduced to solving a simple quadratic programming problem (QPP) or a finite sequence of such problems. A new algorithm is given for solving the QPP, and computational results are included.

---

[2] K.J. Arrow, L. Hurwicz, and H. Uzawa, Studies in Linear and Nonlinear Programming, Stanford University Press, Stanford, California, (1958).

* This paper will appear in the SIAM Journal on Control, Sept. 1965.

Many algorithms are available in the literature for solving the minimum time problem [1-3] as well as quadratic programming problems [4, 5], but in most cases these algorithms cannot be used for computing optimal controls in a feedback mode because of the large computation times involved. The primary justification for further consideration of these problems is in decreasing the computation time.

PAM sample data system

Let X be a discrete, time invariant system described by the linear vector difference equation

$$\underline{x}_{k+1} = A\,\underline{x}_k + \underline{b}\,u_{k+1}, \qquad k = 0, 1, \ldots, \tag{1}$$

where $\underline{x}_k \in E^n$ is the state of X at time k, A is an $n \times n$ constant nonsingular matrix, $\underline{b} \in E^n$ is a constant vector, and the scalar input u_k is constrained in magnitude by $-1 \leq u_k \leq 1$. Note that no difficulties are encountered for the multiple input case, i.e., where $u_k \in E^r$, $r \leq n$. We assume a single input in order to symplify notation. For a given input sequence $u_1, u_2, \ldots, u_k = \underline{u}_k$, it is assumed that the energy supplied to the system X is given by

$$J(\underline{u}_k) = \sum_{i=1}^{k} u_i^2$$

Given $\underline{x}_0$, the initial state of X, we can iterate Eq. 1 and express $\underline{x}_k$ and the control sequence $\{u_j : j=1, 2, \ldots, k\}$:

$$\underline{x}_k(\underline{x}_0, \underline{u}_k) = A^k(\underline{x}_0 + \sum_{i=1}^{k} A^{-i}\,\underline{b}\,u_i). \tag{2}$$

---

[1] Wing, J. and Desoer, C.A., "The Multiple Input Minimal Time Regulator Problem", IEEE Trans. on Automatic Control, Vol. AC-8, No. 2, pp. 125-136, 1963.

[2] Whalen, B.H., "On Optimal Control and Linear Programming," IRE Trans. on Automatic Control (Correspondence), Vol. AC-7, pp. 45-46, 1962.

[3] Tou, J.T., "Optimum Control of Discrete Systems Subject to Saturation," IEEE Trans. on Automatic Control, pp. 88-89, Jan. 1964.

[4] Beale, E.M.L., " On Quadratic Programming", Naval Res. Log. Quart., 6, pp. 227-244, Sept. 1959.

[5] Wolfe, P., "The Simplex Method for Quadratic Programming," Econometrica, 27, pp. 382-398, 1959.

Definition: A control sequence $\underline{u}_k$, of length k, is said to belong to the constraint set Ω if $|u_i| \leq 1$, $i = 1, 2, \ldots, k$.

Problem I. Minimum Energy Problem

Given the initial state $\underline{x}_0$ of X, an integer $N \epsilon \{1, 2, \ldots\}$, and a desired terminal state $\underline{z}_N \epsilon E^n$, find a control sequence $\underline{u}_N^0$ which minimizes

$$J(\underline{u}_N) = \sum_{i=1}^{N} u_i^2 \tag{3}$$

subject to the n + N constraints

$$\underline{x}_N(\underline{x}_0, \underline{u}_N) = \underline{z}_N, \tag{4}$$

$$\underline{u}_N \epsilon \Omega . \tag{5}$$

Let $\underline{r}_i = A^{-i}\underline{b}$, $i=1, 2, \ldots, N$, then Eq. 4 can be written as

$$\sum_{i=1}^{N} \underline{r}_i u_i = (A^{-N} \underline{x}_N - \underline{x}_0) \stackrel{\Delta}{=} \underline{v}_N .$$

Since $\underline{x}_0$, N, and $\underline{z}_N$ are given, $\underline{v}_N$ is known and hence the minimum energy problem can be restated in the following equivalent form :

minimize $$J(\underline{u}_N) \tag{6}$$

subject to $$\sum_{i=1}^{N} \underline{r}_i u_i = \underline{v}_N , \tag{7}$$

$$\underline{u}_N \epsilon \Omega . \tag{8}$$

It is easily shown that if the constraints, Eqs. 7 and 8 admit at least one solution, then the minimum energy problem has a solution. More precisely, if we let $R_N = \{ \sum_{i=1}^{N} \underline{r}_i u_i : \underline{u}_N \epsilon \Omega \}$, then a solution to Problem I exists iff $\underline{v}_N \epsilon R_N$.

Problem II. Intercepting a Moving Target in Minimum Time

Let $\underline{z}_N \in E^n$, N=1, 2, . . . , represent a moving target at time N. Given the initial state $\underline{x}_0$ of X, and the target $\{\underline{z}_N : N=1, 2, \ldots\}$, find a control sequence $\underline{u}_N \in \Omega$, of minimum length N, such that $\underline{x}_N(\underline{x}_0, \underline{u}_N) = \underline{z}_N$. As in Problem I, define $\underline{r}_i = A^{-i}\, \underline{b}$ and $\underline{v}_i = (A^{-i}\, \underline{x}_i - \underline{x}_0)$, i=1, 2, ... An equivalent formulation is: find the smallest integer N such that

$$\sum_{i=1}^{N} \underline{r}_i\, u_i = \underline{v}_N ,$$

$$\underline{u}_N \in \Omega .$$

Note that a solution to this problem exists iff $\underline{v}_N \in R_N$ for some finite integer N. Clearly, a solution to Problem II can be obtained by finding the smallest N for which Problem I has a solution.

Problem III. Minimum Energy Plus Time

Given $\underline{x}_0$ and the target state $\underline{z}_N \in E^n$ at time N, n=1, 2, . . . , find a control sequence $\underline{u}_N^0$ which minimizes

$$Q(\underline{u}_N, N) = \alpha N + \sum_{i=1}^{N} u_i^2, \qquad \alpha > 0$$

subject to the constraints $\underline{x}_N(\underline{x}_0, \underline{u}_N) = \underline{z}_N$ and $\underline{u}_N \in \Omega$. As before, this problem is reducible to the following: find a control $\underline{u}_N^0$ which minimizes $Q(\underline{u}_N, N)$, subject to the constraints $\underline{v}_N = \sum_{i=1}^{N} \underline{r}_i\, \underline{u}_N \in \Omega$. It is easily shown that this problem has a solution iff $\underline{v}_N \in R_N$ for some N, and, furthermore, that a solution can be obtained by solving, sequentially, Problem I starting with N=1, N=2, etc.

Problem IV. Intercepting a Moving Target on a Subspace

In some cases it is not required that the system state agree with the target state in all coordinates; e.g., it may be required that the system state and the target agree in position and velocity. In general it may be required that a linear function of the system state agree with a linear function of the target state. We can formulate this problem as follows. Let H be an $s \times n$ matrix and let $\{\underline{z}_N : N=1, 2, \ldots\}$ be the target. The problem is to find a control sequence $\underline{u}_N \in \Omega$, of minimum length N, such that

$$H\,\underline{x}_N(\underline{x}_0, \underline{u}_N) = H\,A^N \underline{x}_0 + \sum_{i=1}^{N} H\,A^{N-i}\,\underline{b}\,u_i = H\,\underline{z}_N .$$

Defining,

$$\hat{\underline{r}}_i = H\,A^{i-1}\underline{b} \qquad\qquad i=1, 2, \ldots$$

$$\hat{\underline{v}}_i = H(\underline{z}_i - A^i \underline{x}_0).$$

An equivalent formulation is to find the smallest integer N for which

$$\sum_{i=1}^{N} \hat{\underline{r}}_i\, u_{N+1-i} = \hat{\underline{v}}_N$$

$$|u_{N+1-i}| \leq 1, \quad i=1, 2, \ldots, N.$$

It is clear that Problem IV is of the same form as Problem II, and, therefore, a solution can be obtained as described in Problem II. Note, however, that if the target state and the system state need only agree in position, then the problem reduces to the time optimal control of a first order system.

<u>Analysis</u> We have shown that solutions to the four problems stated above can be obtained by solving sequentially Problem I. Therefore, we shall limit our discussion to this problem, and state briefly how a solution is obtained. For a proof of the theorems and lemmas which follow, the reader is referred to reference [6].

<u>Lemma 1:</u> If $\underline{u}_N^0$ is the solution to the minimum energy problem (Problem I), then there exists a constant vector $\underline{c} \in E^n$ such that $u_i^0 = \text{sat} < \underline{r}_i, \underline{c} >$, $i=1, 2, \ldots, N.$*

We have previously remarked that a solution to the minimum energy problem exists iff $\underline{v}_N \in R_N$, consequently, using Lemma 1 we have:

---

[6] Canon, M.D. and Eaton, J.H., "A New Algorithm for a Class of Quadratic Programming Problems with Application to Control," presented at First International Conference on Programming and Control, to appear SIAM Journal on Control, Sept. 1965.

* sat $y \triangleq y$ if $|y| \leq 1$, $\triangleq y/|y|$ if $|y| > 1$

Theorem 1: Each point $\underline{v}_N \in R_N$ can be represented in the form

$$\underline{v}_N = \sum_{i=1}^{N} \underline{r}_i \text{ sat} < \underline{r}_i, \underline{c} > \overset{\Delta}{=} \underline{f}_N(\underline{c})$$

for some vector $\underline{c} \in E^n$.

The minimum energy problem has been reduced to finding a vector $\underline{c} \in E^n$ such that $\underline{v}_N = \underline{f}_N(\underline{c})$, i.e., inverting $\underline{f}_N$. Note that $\underline{f}_N$ maps E^n onto R_N, however, the mapping is not one to one. We next show that it is possible to restrict the domain of $\underline{f}_N$, to a subset of E^n, in such a manner as to make $\underline{f}_N$ a bijective bicontinuous function. Since we wish to find an algorithm for determining $\underline{c}$ given $\underline{v}_N$ (or determining if a solution exists, i.e., if $\underline{v}_N \in R_N$) the continuity of $\underline{f}_N^{-1}$ is of major importance.

Definition: For each $\underline{c} \in E^n$, let $I_N(\underline{c})$ $\{1, 2, \ldots, N\}$ be an index set such that if $i \in I_N(\underline{c})$ then $| < \underline{r}_i, \underline{c} > | \leq 1$; $\bar{I}_N(\underline{c})$ denotes the complement of this set relative to $\{1, 2, \ldots, N\}$. Using this notation $\underline{f}_N(\underline{c})$ can be written as

$$\underline{f}_N(\underline{c}) = \sum_{i \in I_N(\underline{c})} \underline{r}_i \text{ sat} < \underline{r}_i, \underline{c} > + \sum_{i \in \bar{I}_N(\underline{c})} \underline{r}_i < \underline{r}_i, \underline{c} >$$

Definition: Let $C_N \subset E^n$ be the set of all points $\underline{c} \in E^n$ for which the vectors $\{\underline{r}_i : i \in \bar{I}_N(\underline{c})\}$ span E^n.

It is now possible to prove:

Theorem 2: The mapping $\underline{f}_N : C_N \longrightarrow R_N$ is a homeomorphism.

Using the continuity of $\underline{f}_N^{-1}$, a finite step algorithm has been developed for inverting $\underline{f}_N$ [6]. If there is no solution to $\underline{v}_N = \underline{f}_N(\underline{c})$, then the algorithm terminates in a finite number of steps. Thus, it can be determined in a finite number of steps whether $\underline{v}_N \in R_N$, and, as a result, solutions to Problems II through IV can be obtained in a finite number of steps.

To test the computational efficiency of this algorithm time optimal controls of length 20 sampling periods or less were computed for a fourth order system (i.e., $\underline{r}_i \in E^4$). Using an IBM 7090 computer the maximum computation time was 0.4 seconds. In solving the minimum time problem for an optimal control of length 20 sampling periods the minimum energy problem is solved 20 times. Thus for N = 20, Problem I can be solved in approximately 0.07 seconds.

A NEW APPROACH TO THE SOLUTION OF QUADRATIC PROGRAMMING PROBLEMS

NASA Grant NsG-354 (Supplement 2)
M. Canon (Prof. E. Polak)

All the algorithms presently available for solving quadratic programming problems (QPP) share one common feature, viz., at each step of the algorithm the boundary conditions are satisfied and the value of the cost function is reduced. In this note we show how the QPP can be reduced for solving a set of simultaneous nonlinear equations. A finite-step algorithm has been developed for solving these equations.

The Quadratic Programming Problem

Find n real variables $x_1^0, x_2^0, \ldots, x_n^0$ (representing an n-vector $\underline{x}^0$) which minimize

$$J(\underline{x}) = \langle \underline{x}, Q\underline{x} \rangle + \langle \underline{x}, \underline{d} \rangle \tag{1}$$

subject to the constraints

$$A\,\underline{x} = \underline{b} \tag{2}$$

$$x_i \geq 0, \quad i=1, 2, \ldots, n \tag{3}$$

Here Q is a symmetric, positive semi-definite $n \times n$ matrix, A is an $m \times n$ matrix of full rank, $\underline{d} \in E^n$ and $\underline{b} \in E^m$ are constant vectors. Let N(Q) and N(A) denote, respectively, the null space of the operators Q and A. It is assumed that $N(Q) \cap N(A) = \{\underline{0}\}$, the zero vector.

Definition: An n-vector $\underline{x}$ is said to belong to the constraint set Ω, if $x_i \geq 0$, $i=1, 2, \ldots, n$.

Definition: Let $\mathcal{A}$ be the image of the constraint set Ω under the linear transformation A, i.e., $\mathcal{A} = \{A\,\underline{x} : \underline{x} \in \Omega\}$.

In a straightforward manner one can prove the following:

Lemma 1: If $\underline{b} \in \mathcal{A}$, then the QPP has a solution; furthermore, if $N(Q) \cap N(A) = \{\underline{0}\}$, then the solution is unique.

Necessary and Sufficient Conditions for Optimality

Following Kuhn and Tucker [1] we introduce a scalar valued function $H(\underline{c}, \underline{x})$ defined by

$$H(\underline{c}, \underline{x}) = \langle c, A x \rangle - J(\underline{x}),$$

where $\underline{c} \in E^m$, $J(\underline{x})$ and A are defined above. Let $(H_{\underline{x}^0})_i$, $i=1, 2, \ldots, n$ denote the $\underline{i}$th component of the partial derivative of H with respect to $\underline{x}$, evaluated at $\underline{x} = \underline{x}^0$ and $\underline{c} = \underline{c}^0$. After a slight modification of the Kuhn and Tucker theorem [1], it is possible to prove Theorem 1.

Theorem 1: A necessary and sufficient condition for $\underline{x}^0$ to be a solution of the QPP is the existence of a vector $\underline{c}^0 \in E^m$ such that

$$\text{(i)} \quad \text{If } x_i^0 = 0, \text{ then } (H_{\underline{x}^0})_i \leq 0 \tag{4a}$$

$$\text{(ii)} \quad \text{If } x_i^0 > 0, \text{ then } (H_{\underline{x}^0})_i = 0 \tag{4b}$$

It turns out that Eq. 2 implies $H(\underline{c}^0, \underline{x}^0) = \max\limits_{\underline{x} \in \Omega} H(\underline{c}^0, \underline{x})$.

The Vector-Valued Function f

For ease in explanation, let us assume that Q is positive definite. A slight modification of the following argument is necessary under the more general assumption $N(Q) \cap N(A) = \{\underline{0}\}$.

Let $\underline{g}$ be the vector function mapping E^m into Ω defined as follows: to each $\underline{c}^0 \in E^m$, $\underline{g}(\underline{c}^0)$ is that point in Ω which satisfies Eq. 4, i.e.,

$$H(\underline{c}^0, \underline{g}(\underline{c}^0)) = \max_{\underline{x} \in \Omega} H(\underline{c}^0, \underline{x}). \tag{5}$$

It is easily shown that $\underline{g}$ is a function, i.e., to each $\underline{c}^0 \in E^m$ there is one, and only one, $\underline{x}^0 \in \Omega$ satisfying Eq. 3. One can now associate to each $\underline{c} \in E^m$ a point $\underline{a} \in \mathcal{A}$ by the composite function $A \circ \underline{g} = \underline{f}$, where A is the matrix in Eq. 2. By Lemma 1, a solution to the QPP exists if the linear equality constraint, Eq. 2, is replaced by $A \underline{x} = \underline{a}$, for all $\underline{a} \in \mathcal{A}$. Consequently, it follows from Theorem 1, that for every $\underline{a} \in \mathcal{A}$ there exists a $\underline{c} \in E^m$ such that $\underline{f}(\underline{c}) = \underline{a}$; clearly, $\underline{f}$ is continuous. We have proved Theorem 2.

---

[1] H.W. Kuhn and A.W. Tucker, "Nonlinear Programming," Proc Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, 1951, pp. 481-492.

Theorem 2: The mapping $\underline{f}: E^m \rightarrow \mathcal{A}$ is continuous and onto, furthermore, if $\underline{a} = \underline{f}(\underline{c}) = (A \circ g)\,\underline{c}$, then $g(\underline{c})$ is the solution to the QPP:

minimize $\quad J(\underline{x})$

subject to $\quad A\underline{x} = \underline{a}$

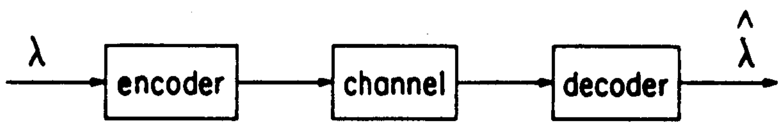
$$\underline{x} \in \Omega ,$$

The QPP has now been reduced to inverting the equation $\underline{f}(\underline{c}) = \underline{b}$. A finite step algorithm has been developed for performing this inversion and a computer program has been written and tested on several problems. Preliminary computational results are favorable.

Unfortunately, we do not yet have a method for proving that a given signal configuration is a local minimum rather than a type of saddle point. A more detailed discussion of these results is expected to appear in the next issue of "Notes on System Theory", (to be published as an ERL report, Spring 1966).

LOWER BOUNDS ON MEAN-SQUARED ERROR

NASA Grant NSG-354 (Supplement 2)
B. Haskell (Prof. D.J. Sakrison)



Consider the single parameter transmission system shown above, where the channel has capacity C bits/sec. The mutual information between the input and output per unit time is less than or equal to C. We wish to maximize this transmitted information while at the same time minimizing the mean squared error,

$$\delta^2 = \iint (\lambda - \hat{\lambda})^2 \, p(\hat{\lambda}/\lambda) \, p(\lambda) \, d\lambda \, d\hat{\lambda} \tag{1}$$

$p(\hat{\lambda}/\lambda)$ is the transition probability density function and $p(\lambda)$ is the a priori density function.

For a given $p(\lambda)$, define the rate distortion function as the minimum mutual information possible if the mean-squared error is less than or equal to D.

$$R(D) = \min_{p(\hat{\lambda}/\lambda)} \iint p(\hat{\lambda}/\lambda) \, p(\lambda) \, \log_2 \frac{p(\hat{\lambda}/\lambda)}{p(\lambda)} \, d\lambda \, d\hat{\lambda}, \tag{2}$$

where $p(\hat{\lambda}/\lambda)$ is varied over the class defined by

$$\delta^2 = \iint p(\hat{\lambda} - \lambda)^2 \, p(\lambda/\lambda) \, p(\lambda) \, d\lambda \, d\hat{\lambda} \leq D. \tag{3}$$

If T is the time devoted to the transmission of one parameter, then

$$R(D) \leq CT. \qquad (4)$$

An equivalent statement is that if $CT = R(D)$, then $\delta^2 \geq D$. (5)

Similarly, if $p(\hat{\lambda}/\lambda)$ is varied over a subclass of that defined by inequality (3) and $R_1(D)$ is obtained, then when using a system in this class, $CT = R_1(D)$ implies that $\delta^2 \geq D$. (6)

Consider the a priori density function

$$p(\lambda) = \begin{cases} \frac{1}{2} & |\lambda| \leq 1 \\ 0 & \text{elsewhere} \end{cases} \qquad (7)$$

Using statement (6) we can show that for a PCM system [1]

$$\delta^2 \geq \frac{1}{3}\, 2^{-2CT}, \qquad (8)$$

and we can find a lower bound on the mean-squared error when using PAM, PPM, or FSK. By means of a lower bound on R(D), derived by Shannon, we can also show that for any system [2]

$$\delta^2 \geq \frac{2}{\pi e}\, 2^{-2CT} \qquad (9)$$

These are plotted in Fig. 1.

The PCM lower bound and the absolute lower bound have been derived geometrically for the case of gaussian white noise. It is shown that the bounds can be generalized. (These results are presented in detail in the Master's II report by B.G. Haskell, "Pulse Modulation.")

---

[1] A similar result is shown by A. J. Viterbi, "Maximum SNR for Digital Communications," IEEE Trans. on Communications Systems Vol. CS-12, No. 1, March, 1964.

[2] This result is also derived in Wozencraft and Jacobs, Principles of Communication Engineering, to be published, John Wiley and Sons, Inc.

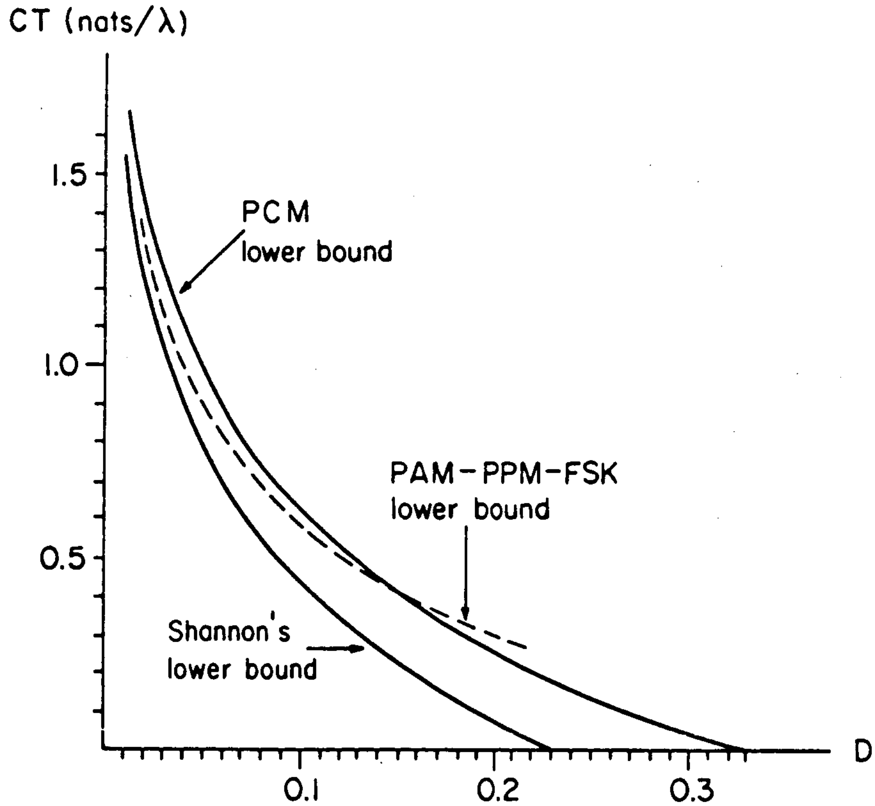


Fig. 1. Lower bounds on δ^2.